REMARKS

Information Disclosure Statement

An Information Disclosure Statement (IDS) is being filed concurrently herewith. Entry of the IDS is respectfully requested.

Claim Rejection under 35 U.S.C. §112, First Paragraph

The Examiner has rejected Claims 1-18 under U.S.C.§112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the claimed invention. In particular, the Examiner states that Claims 1-11 were amended with the following limitation: "in a manner free of predetermined association of patterns with respective clusters," and that this limitation is new matter.

In substitution of the objected claim language, Applicants now amended Claims 1 and 11 with the following limitation: "wherein prior knowledge of the datapoints to be clustered in not necessary." This limitation describes that the methods of the claimed invention is performed in an unsupervised learning fashion. Support for this amendment can be found throughout the specification, and in particular, on specification page 8, lines 1-13; and page 23, lines 4-11 as originally filed. Specifically, page 23, lines 4-11 of the specification as filed states:

> Yeast Cell Cycle: GENECLUSTER™ was tested on a published dataset, to determine whether it could automatically expose known patterns *without using prior knowledge.* For this purpose, data was used from a recent study of Cho, R. *et al.* (1998) *Molecular Cell* 2, 65-73. In the study, the researchers synchronized *S. cerevisiae* in G1, released the cells, and collected RNA at 10 min intervals over two cell cycles (160 min). Expression levels of 6,218 yeast ORFs were measured using oligonucleotide arrays. From the set of genes passing a variation filter, the authors used visual inspection to identify 416 genes showing peaks of expression in early G1, late G1, S, G2 or M phase. *Emphasis added.*

It is clear that support for this limitation exists. No new matter is added. This Amendment is being made to expedite prosecution, and not for reasons related to patentability.

Claim Rejection Under 35 U.S.C. §103(a)

The Examiner rejected Claims 1-18 under 35 U.S.C. §103(a) as being unpatentable over Mack, in view of Mangiameli and Kohonen. The Examiner states that Mack discloses methods

of cluster analysis for gene expression monitoring. In particular, the Examiner states that Mack's methods comprise receiving gene expression values of datapoints, clustering the datapoints, and providing output display indicating the cluster of the datapoints. The Examiner further states that while Mack does not explicitly disclose clustering using SOMs, it does suggest using alternative statistical methods. The Examiner states that Mangiameli applied SOM and seven hierarchical methods to 252 messy data sets and found that SOMs are significantly superior in both robustness and accuracy to other clustering methods. The Examiner asserts that one of ordinary skill in the art would have been motivated to modify the method of Mack to use SOMs as suggested and taught by Mangiameli and Kohonen for the cluster analysis of gene expression data to achieve its superiority in accuracy and robustness.

Applicants respectfully disagree. Applicants are submitting herewith a Declaration under 37 C.F.R. §1.132 by Dr. Gabriel Kreiman. In paragraph 6, the Declarant states that although SOMs have been used for quite some time now, the application of this tool to the study of gene expression data is novel and has unexpected, surprising results. Furthermore, in paragraph 7 of the attached §1.132 Declaration, the Declarant states that Mack only generally refers to clustering methods, and does not suggest using an unsupervised clustering method. The Declarant further quotes a passage of Mack where Mack only refers to very general cluster books. The Declarant further declares that numerous clustering methods and algorithms have been published, and each may serve different purposes. Mack applies a cluster analysis in a unilateral, class inclusion/exclusion (of a certain class) approach. Mack, however, provides no algorithm or specific description of how the clustering analysis is performed. Mack, a reference that only generally refers to clustering, is not specific enough to make the claimed invention, specifically clustering gene expression data using a self organizing map, obvious. More importantly, one of skill in the art, such as the Declarant, would not read the generalizations in Mack, in combination with the other cited references, and come to realize the specific direction of the claimed invention, namely to use an unsupervised clustering algorithm of SOM to analyze gene expression data.

In addition, according to paragraph 8 of the attached §1.132 Declaration, the Declarant would not have combined Mack with Mangiameli and Kohonen. Mack describes mapping regulatory relationship among genes by using supervised clustering where the number of clusters as well as their composition are determined by the scientist beforehand. The Declarant states,

for example, that the supervised clustering, as generally described in Mack, refers to the need for the user to input information into a computer before the algorithm can be used to determine the results. According to the Declarant, supervised clustering, like that described in Mack, is one that involves the predetermination of the number of clusters as well as their composition; in contrast, the claimed invention of Tamayo *et al.* uses unsupervised clustering, which generally does not require user input, and the algorithm can analyze the data independently. The Declarant states that an important distinction between Mack and the claimed invention is that the claimed invention does not require prior knowledge of the data, whereas Mack does. In fact, the Declarant cites the following passage in Mack: "in some embodiments, models are built by incorporating expression data and current knowledge about the regulation of specific genes." Column 27. Additionally, the Declarant did not find an example in Mack in which unsupervised learning was used. Mangiameli and Kohonen describe the use of SOM, an algorithm utilized in unsupervised learning. As such, the Declarant would not have combined Mack which describes supervised clustering with Mangiameli and Kohonen which describe the use of SOM, an algorithm utilized in unsupervised learning, to analyze gene expression data.

According to Mack, it uses cluster analysis for determining whether mutations in up-stream regulatory genes exist by monitoring down-stream gene expression. In particular, Mack provides a method for determining whether a down-stream gene expression indicates a p53 mutation or not. The methods in Mack determine whether or not a gene expression falls into a known, pre-determined category. Contrary to this binary (inclusion/exclusion) classification for known classes, the claimed invention advantageously, in an unsupervised learning fashion, classifies gene expression data into *multiple* classes. Importantly, the claimed invention *does not require any information of known classes*. The claimed invention can classify gene expression data into unknown classes, redefine classes, or rediscover classes.

Combining Mack with Mangiameli would be also improper because one of ordinary skill in the art would not use SOMs, an accurate and robust neural network for messy empirical data as described in Mangiameli, for data requiring only a simple, parametric analysis, like that found in Mack.

Assuming *arguendo* that Mack and Mangiameli were combined, the present invention would not result. Instead one would be motivated to use Mack in a subsequent stage after processing messy empirical data. For example, once the claimed invention is applied to gene

expression data to determine what classes exists, including previously unknown classes, then the methods described in Mack could be used to assign a sample to one of those classifications.

Additionally, the Examiner describes an embodiment of Mack that is a method in which genes are regulated by a gene, p53, (See Column 29) and the method comprises taking samples from such cells containing the gene mutation and analyzing the expression level, and building a causal model. (See Fig. 2). The Examiner also states that it is clear that the clustering and categorizing step has nothing to do with the prior knowledge of the samples being a p53 mutation or not. According to the Examiner, only after cluster analysis is done, by comparing the clusters of p53 mutation and wilde type, a model is obtained for genes regulated by p53. The Examiner states that Mack does not require prior knowledge of any relationship of the genes tested from p53.

According to the Declarant in paragraph 10 of the Declaration, this experimental section has been interpreted to mean that at some point the user will select a predetermined number of clusters and assume some knowledge about the methodology. The Declarant also declares that although Mack does set out to determine the relationship of genes, certain knowledge is assumed and used to determine this relationship.

Even if a prima facie case of obviousness is established, it is rebutted for the following reasons: the claimed invention is surprising and unexpected, and fulfills certain long-felt needs. According to the Declarant (paragraph 11 of the attached Declaration), combination of the cited references leave the following needs: (i) the need for more sophisticated tools (e.g. accurate assessment of statistical significance of results, different types of boundaries for clusters, etc.), (ii) the need to analyze different aspects of the data (e.g. time series data sets, comparison of different conditions, etc.), and (iii) the need for unsupervised clustering tools. To this end, the Declarant states that the claimed invention has fulfilled these long felt needs and has provided the Bioinformatics and Genetics communities with a very valuable and novel tool to analyze gene expression data, a tool not described by the cited references.

Thus no combination of the prior or cited art makes obvious the present invention as now claimed. To highlight the foregoing distinctions over the prior art, base claim 1 as now amended recites:

using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters, **wherein prior knowledge of the datapoints to be clustered in not necessary**. *Emphasis added.*

Claim 11, the other independent claim, has now been amended to recite similar language. The dependent claims inherent the distinguishing claim language of the base claims. Accordingly, Applicants believe that the rejection of Claims 1-18 under 35 U.S.C. §103 is overcome. Reconsideration of Claims 1-18 is respectfully requested.

CONCLUSION

In view of the above amendments and remarks, it is believed that all claims (Claims 1-18) are in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned at (978) 341-0036.

Respectfully submitted,
HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____
Antoinette G. Giugliano
Registration No. 42,582
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133
Dated: 7/8/03

MARKED UP VERSION OF AMENDMENTS

<u>Claim Amendments Under 37 C.F.R. § 1.121(c)(1)(ii)</u>

1.   (Twice Amended) In a computer system, a method for clustering a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises:

    a)   receiving the gene expression values of the datapoints;

    b)   using a self organizing map, clustering the datapoints such that the datapoints that exhibit similar patterns are clustered together into respective clusters [in a manner free of predetermined association of patterns with respective clusters] <u>, wherein prior knowledge of the datapoints to be clustered in not necessary</u>; and

    c)   providing an output indicating the clusters of the datapoints.

11.   (Twice Amended) In a computer system, a method for grouping a plurality of datapoints, wherein each datapoint is a series of gene expression values, wherein the method comprises:

    a)   receiving gene expression values of the datapoints;

    b)   filtering out any datapoints that exhibit an insignificant change in the gene expression value, such that working datapoints remain;

    c)   normalizing the gene expression value of the working datapoints;

    d)   using a self organizing map, grouping the working datapoints such that the datapoints that exhibit similar patterns are grouped together into respective clusters [in a manner free of predetermined association of patterns with respective clusters] <u>, wherein prior knowledge of the datapoints to be clustered in not necessary</u>; and

    e)   providing an output indicating the groups of the datapoints.